

# 全国第六届研究生数学建模竞赛



## 题 目 A 题 我国就业人数或城镇登记失业率的数学建模

### 摘 要：

本文针对就业问题，运用多元线性回归、VAR 模型以及面板数据处理等方法，分别建立全国和各地区城镇就业人数与其主要影响因素之间的函数关系，并利用 VAR 模型对我国 2009 年及 2010 年上半年的就业前景进行仿真。本文的亮点在于：第一，通过两种方法从不同角度寻找出了影响就业的两套主要因素；第二，在分地区进一步细化模型后，利用面板数据，建立了个体固定效应模型；第三，运用制度经济学的分析方法对提高城镇就业人口数提出了若干独到的意见和建议。

针对问题 1，一方面，根据灰色关联度原则和模糊数学中最大最小贴近度准则，建立影响就业的主要因素模型，运用 MATLAB 软件编程求解，得出影响就业的主要因素为全国总人口、劳动生产率、全国城镇单位平均工资、三产产值和 CPI；另一方面，利用逐步回归法，运用 Eviews 软件对影响就业的 7 个指标进行回归，得出影响就业的主要因素为全国总人口、GDP、三产产值、CPI 和 M2。

针对问题 2，利用逐步回归方法，采用 Eviews 软件编程，建立了全国城镇就业人数的两个多元回归模型。一个模型的结论为：城镇就业人数与全国总人口、GDP 和 CPI 成负相关，与三产产值和 M2 成正相关；另一个模型的结论为：城镇就业人数与劳动生产率和 CPI 成负相关，与全国城镇单位平均工资成正相关，表现为其每增长 1 个百分点，可以拉动城镇就业人数增长 0.4097 个百分点。

针对问题 3，针对面板数据建立了个体固定效应模型，利用 Eviews 软件编程求解，得出全国各地城镇就业人数与劳动生产率、三产产值和 CPI 成负相关，与 GDP 成正相关，且 GDP 每增加 1 个百分点，城镇就业人数将被拉动增长 0.5739 个百分点。

针对问题 4，在问题 2 的基础上建立 2 个 VAR 模型，通过对样本期内的数据拟合，发现两个模型精度都相当高，在两个模型比较后，得出 VAR2 模型在预测精度上优于 VAR1，最后运用 VAR2 模型预测出 2009 年全国城镇就业人数为 31304 万人，比上年增长 1094 万人，与中央预期新增就业 900 万人以上的目标一致。

**关键字：**灰色关联度 VAR 模型 个体固定效应模型 制度分析

**参赛队号** 1164601

**队员姓名** 丁井国 俞佳佳 胡钢

**参赛密码** \_\_\_\_\_

**(由组委会填写)**

# 目录

1 问题重述.....	3
2 模型假设.....	3
3 符号说明.....	4
4 问题分析.....	4
4.1 问题 1 分析.....	4
4.2 问题 2 分析.....	4
4.3 问题 3 分析.....	4
4.4 问题 4 分析.....	5
4.5 问题 5 分析.....	5
5 模型建立与求解.....	5
5.1 问题 1 模型与求解.....	5
5.1.1 理论分析.....	5
5.1.2 问题 1 模型 I .....	6
5.1.3 问题 1 模型 II .....	7
5.1.4 问题 1 模型 II 求解——最大值和最小值贴近度的计算 .....	9
5.1.5 问题 1 模型 II 求解——灰关联度计算 .....	10
5.1.6 问题 1 结论.....	11
5.2 问题 2 模型与求解.....	11
5.2.1 问题 2 模型 I 的建立.....	11
5.2.2 问题 2 模型 II 的建立.....	13
5.2.3 问题 2 模型 II 的结论.....	18
5.3 问题 3 模型.....	18
5.3.1 问题 3 数据的检验.....	18
5.3.2 问题 3 模型选择.....	19
5.3.3 模型优化.....	21
5.3.4 问题 3 结论.....	22
5.4 问题 4 仿真（预测）结果.....	23
5.4.1 预测方案 1.....	23
5.4.2 预测方案 2.....	24
5.4.3 两个VAR模型优劣比较.....	26
5.5 政策建议.....	27
5.5.1 模型结果及原因分析.....	27
5.5.2 基于制度经济学角度的政策建议.....	28
6 模型优缺点.....	29
7 参考文献.....	29

## 1 问题重述

失业、经济增长和通货膨胀为宏观经济中特别重要的三个指标，就业（或者失业）是社会、国民经济中极其重要的问题。按照已有研究，就业可以定义为三个月内有稳定的收入或与用人单位有劳动聘用关系。失业的统计方法各国差异较大，我国采用城镇登记失业率，是指城镇登记失业人数同城镇从业人数与城镇登记失业人数之和的比。其中，城镇登记失业人员是指有非农业户口，在一定的劳动年龄内，有劳动能力，无业而要求就业，并在当地就业服务机构进行求职登记的人员。

从经济学的角度，影响就业（或者失业）的因素很多。从宏观层面上，消费、投资、政府购买和进出口都是重要的因素；而从中观层面，不同地区、不同产业也会表现出不同的特征。当然，中央政府调整宏观经济政策（包括财政政策和货币政策），以及对不同地区和不同产业实行不同的扶持政策都会对就业产生巨大的影响。

就我国的现实情况，2008 年我国经济社会经受了历史罕见的考验，GDP 依然保持 9% 以上平稳较快增长，城镇新增就业 1113 万人，城镇登记失业率为 4.2%。2009 年我国就业面临更大的挑战，一是国际金融危机导致国际市场需求难以在短期内复苏；二是今年我国经济增长速下滑；三是国内消费需求乏力；四是一些行业产能过剩与市场预期不确定导致企业投资不足，所以就业形势十分严峻。

为此，中央政府从 08 年 10 月开始实施了 40000 亿元的投资计划，确定了十大产业振兴计划，采取扩大国内消费需求的措施，提高对外开放水平以增加出口。同时，中央财政拟投入 420 亿元资金实施积极的就业政策。09 年我国在就业方面的目标：城镇新增就业 900 万人以上，城镇登记失业率控制在 4.6% 以内。

利用近年来我国有关的统计数据并结合一年多来我国国民经济的运行数据，就我国就业人数或城镇登记失业率研究如下问题。

1. 对有关统计数据进行分析，寻找影响就业的主要因素或指标。
2. 建立城镇就业人数或城镇登记失业率与上述主要因素或指标之间联系的数学模型。
3. 对上述数学模型从包含主要的经济社会指标、分行业、分地区、分就业人群角度，尝试建立比较精确的数学模型。（由于时间限制，建议适度即可）
4. 利用所建立的关于城镇就业人数或城镇登记失业率的数学模型，根据国家的有关决策和规划对 2009 年及 2010 年上半年的我国就业前景进行仿真
5. 根据所建立的数学模型和仿真结果，对提高我国城镇就业人口数或减少城镇登记失业率提出你们的咨询建议。

## 2 模型假设

假设 1：官方统计年鉴所得数据真实可靠，能够反映指标所代表的真实含义。

假设 2：城镇登记失业率可以较为真实反映中国的实际就业情况。

假设 3：在预测年份中，没有发生重大的灾难性事件。

### 3 符号说明

符号	定义
$Y$	城镇就业人数
$X_1$	全国(地区)总人口
$X_2$	劳动生产率(工业增加值/全国就业人员)
$X_3$	城镇单位平均工资
$X_4$	经济增长(GDP)
$X_5$	第三产业产值
$X_6$	消费价格指数(CPI)
$X_7$	货币政策(M2)
$A_i$	各项指标的模糊子集( $i=1\ldots 7$ 时, 指标为 $X_i$ ; $i=8$ 时, 指标为 $Y$ )

### 4 问题分析

本题主要是涉及到宏观、中观经济背景的分析, 要求参赛者具有一定的经济学素养, 同时也要求具备基本的计量经济学知识。该题主要要求参赛者先找出影响就业的主要因素, 建立它们之间联系的数学模型, 并从分地区、行业角度建立更加精确的模型, 最后要求参赛者对未来一段时期的就业前景进行仿真预测, 并提出建议。由此看来, 存在的主要问题有: 1、如何找出影响就业的主要因素? 2、怎样建立模型? 3、如何细化模型并对未来预测? 这些问题的解决离不开计量经济学知识。因此, 笔者认为, 首先可以考虑运用灰色关联度确定这些主要因素; 其次考虑利用计量经济学中的多元线性或非线性回归建立模型; 再次, 对所建立的模型结合未来一段时期内的就业形势进行预测, 并提出意见和建议。

#### 4.1 问题 1 分析

对于问题一, 我们首先从理论上分析可能会影响到就业的因素。在初步确定影响就业的众多因素之后, 再尝试用量化的方法找出哪些是主要因素, 哪些是次要因素。

#### 4.2 问题 2 分析

该问题其实就是假设知道了解释变量和被解释变量, 怎么来确定它们之间的关系。通常我们会想到计量经济学中的很多模型, 比如多元线性回归、非线性回归等模型, 在对就业指标及其影响因素具体分析后, 即可尝试建立相应的回归模型。

#### 4.3 问题 3 分析

问题 3 则是强调从更细的方面来考虑, 旨在期望研究生们能够建立更加精确的模型。这个题目要求“不求全而求专”, 即以社会经济生活的某一侧面展开分

析，寻求影响该领域的就业的主要因素，并建立相应的数学模型。

#### 4.4 问题 4 分析

对于问题 4，我们根据已经建立的模型，对未来一段时期内的就业形势作出预测，在这里，要求根据国家的有关决策和规划，结合产业结构调整等相关的指标，对 2009 年及 2010 年我国的就业情况进行仿真。

#### 4.5 问题 5 分析

问题 5 要求参赛者根据所建立的数学模型的预测效果，结合当前我国实际，提出一些能够改善国家就业压力的可行性措施、建议；当然，要结合相关的经济学理论进行分析。

### 5 模型建立与求解

#### 5.1 问题 1 模型与求解

##### 5.1.1 理论分析

按照西方经济学理论结合中国经济的实际情况，分析就业的主要影响因素有：

1. 劳动力数量（指标：全国总人口数）

劳动力数量决定劳动力的供给。当劳动力数量增加时，劳动力供给会增加，由于工资的刚性作用，供给不会立即发生变化，从而新增加的劳动力会在这一工资水平下加入到就业人员行列，因此在此种状况下就业会有所变化。

2. 劳动生产率（指标：工业增加值 / 全国就业人员）

劳动生产率是指经济活动的投入和产出比率。技术进步是劳动生产率提高的主要因素，并且从短期来看，劳动生产率越高，失业人口会增加，相应的就业人口就会减少，这是因为一部分就业人员被机器所代替。但长期来看，劳动生产率发展会带来经济的增长，也是会有利于增加就业人数的。

3. 工资水平（指标：城镇单位平均工资）

工资水平是用来衡量劳动力价值的价格。工资水平取决于劳动力的供给与需求，是市场化的行为，但在现实中工资水平还受国家政策和工会组织等的影响；工资水平不单是被动地受供求关系的影响，而且也会主动的影响到劳动力的供给与需求。一般来说，就业人数与工资水平呈同向变化的趋势。

4. 经济增长率（指标：GDP）

经济增长是指一个较长的时间跨度上，一个国家人均产出（或人均水平）水平持续增加。完整的增长衡量主要包括有消费、投资、政府购买、进出口等方面，这些主要的经济指标都与就业息息相关，对就业的影响程度是不可忽视的。依照奥肯定律，失业率与经济增长之间存在交替关系，相应的，经济增长与就业之间存在同向变化的趋势问题。

5. 产业结构（指标：第三产业产值）

产业结构是指各产业部门之间以及各产业部门内部的构成，一般的产业构成从三次产业划分法来进行分析。产业结构变化更多的是体现社会不同部门的分工趋向性，不同产业的就业状况不同，因此，产业结构与就业状况联系就十分紧密。

6. 通货膨胀率（指标：CPI）

通货膨胀率是货币超发部分与实际需要的货币量之比，用以反映通货膨胀、货币贬值的程度；而价格指数则是反映价格变动趋势和程度的相对数。依据菲利普斯曲线，通货膨胀率和失业率是此消彼长的关系，后来各派学者亦对这一结论进行分析和论证，也得出在不同经济环境下的不同结论，但是两者之间的存在关系确实一定的。

## 7. 货币政策（指标：M2）

货币政策属于国家宏观调控的重要手段之一，主要表现在于调控市场货币数量来达到紧缩或者放松银根进而影响经济。起其实质就是金融对经济的作用，货币政策直接关系到整个市场经济的活跃程度，因此，对于就业的影响是需要考虑的。另一方面，货币政策是经济手段但是包含有政治体制因素在其中，对于市场作用程度也会影响到就业状况。

由以上的分析，我们决定选取全国总人口、劳动生产率、全国城镇单位平均工资、GDP、第三产业产值、CPI、M2 七个指标作为影响因素因素，相关数据见附件 1.1 所示：

对于问题 1, 我们得出两个模型。我们将在 5.4.3 节来讨论这两个模型的优劣性。

### 5.1.2 问题 1 模型 I

我们采用逐步回归的方法来求  $Y$  与  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$  之间的回归模型。

多元线性回归模型为：

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6 + \alpha_7 X_7 + \varepsilon$$

回归过程和结果如表 1 所示(表中标识为蓝色单元格是逐步回归每步下的最优模型，同表 7)：

表 1  $Y$  模型逐步回归预测结果

模型	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$R^2$	$\overline{R}^2$
$Y = f(X_5)$					0.13 (17.77)			0.95	0.95
$Y = f(X_5, X_1)$	0.36 (8.23)				0.07 (8.81)			0.99	0.99
$Y = f(X_5, X_2)$		-1.16 (-2.06)			0.28 (3.74)			0.96	0.95
$Y = f(X_5, X_3)$			-1.04 (-1.13)		0.38 (1.69)			0.95	0.95
$Y = f(X_5, X_4)$				-0.10 (-1.97)	0.37 (2.98)			0.96	0.95
$Y = f(X_5, X_6)$					0.12 (19.00)	-83.66 (-2.68)		0.96	0.96
$Y = f(X_5, X_7)$					0.13 (0.85)		0	0.95	0.94
$Y = f(X_5, X_1, X_2)$	0.34 (11.74)	-0.86 (-4.68)			0.18 (7.26)			1.00	1.00
$Y = f(X_5, X_1, X_3)$	0.39 (8.05)		0.55 (1.20)		-0.07 (-0.60)			0.99	0.99
$Y = f(X_5, X_1, X_4)$	0.35			-0.08	0.27			1.00	1.00

	(13.08)			(-5.40)	(7.18)				
$Y = f(X_5, X_1, X_6)$	0.32 (8.19)				0.07 (10.85)	-40.95 (-2.78)		0.99	0.99
$Y = f(X_5, X_1, X_7)$	0.41 (12.26)				-0.14 (-2.75)		0.05 (4.07)	1.00	0.99
$Y = f(X_5, X_1, X_4, X_2)$	0.36 (13.13)	1.08 (1.14)		-0.18 (-2.09)	0.36 (4.14)			1.00	1.00
$Y = f(X_5, X_1, X_4, X_3)$	0.36 (12.21)		0.30 (1.05)	-0.08 (-5.16)	0.19 (2.22)			1.00	1.00
$Y = f(X_5, X_1, X_4, X_6)$	0.33 (13.09)			-0.07 (-4.81)	0.24 (6.81)	-22.83 (-2.27)		1.00	1.00
$Y = f(X_5, X_1, X_4, X_7)$	0.38 (21.79)			-0.07 (-6.52)	0.08 (1.80)		0.03 (5.14)	1.00	1.00
$Y = f(X_5, X_1, X_4, X_7, X_2)$	0.38 (21.58)	-0.60 (-0.87)		-0.01 (-0.15)	0.01 (0.11)		0.04 (4.83)	1.00	1.00
$Y = f(X_5, X_1, X_4, X_7, X_3)$	0.37 (22.5)		-0.39 (-1.99)	-0.06 (-7.04)	0.13 (2.76)		0.04 (5.69)	1.00	1.00
$Y = f(X_5, X_1, X_4, X_7, X_6)$	0.36 (29.27)			-0.05 (-7.84)	0.06 (2.16)	-20.02 (-4.34)	0.03 (7.35)	1.00	1.00
$Y = f(X_5, X_1, X_4, X_7, X_6, X_2)$	0.36 (28.09)	-0.08 (-0.17)		-0.05 (-1.05)	0.05 (0.85)	-19.79 (-3.97)	0.03 (6.06)	1.00	1.00
$Y = f(X_5, X_1, X_4, X_7, X_6, X_3)$	0.36 (28.08)		0.01 (0.07)	-0.05 (-7.38)	0.06 (1.41)	-20.29 (-3.24)	0.03 (4.76)	1.00	1.00

注：表格的单元格中，上面的是系数，下面的是  $t$  值(下同)。

由表 1 可知，模型中影响  $Y$  的主要因素只保留了 5 个，即： $X_1$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$ 。

### 5.1.3 问题 1 模型 II

在本模型中，我们从两个方面来考虑：一是模糊数学中的最大最小贴近度；二是灰度关联系数。

模糊数学中，贴近度常用来反映两个模糊集合的相近程度，贴近度越接近于 1，则表明两模糊集越接近，否则，两模糊集越远。因此，利用模糊数学中的贴近度指标可以在一定程度上反映出影响某个指标的一批因素的相对重要程度。

#### (1) 模糊子集及隶属度函数<sup>[1]</sup>

对于某一集合  $A$ ，元素  $a$  要么属于  $A$ ，要么不属于  $A$ 。二者必居其一，且仅仅居其一。这是经典集合的特征。对于这一特征，经典集合  $A$  中的元素  $u$  与集合  $A$  的关系可以用一个函数——特征函数来刻画。集合  $A$  的特征函数(也成示性函数)是指：

$$I_A(u) = \begin{cases} 1 & u \in A \\ 0 & u \notin A \end{cases} \quad (1)$$

也就是说，若  $u$  是  $A$  中元素，则其特征函数  $I_A(u)$  的值为 1，若  $u$  不是  $A$  中元素，则特征函数  $I_A(u)$  的值为 0。上述特征函数体现了经典集合论中的“非此

即彼”的二值逻辑关系。

然而在实际工作和生活中，我们常遇到并非都是“非此即彼”的情况，而是介于“是”与“不是”之间，表现出“亦此亦彼”的特征。比如，对于某种商品的质量，由人认为很好，由人认为一般，也有人认为较差。因此，我们很难用好或不好来形容该商品的质量。

为了解决实际工作中的这类问题，我们必须把元素属于集合的概念模糊化，变经典集合的“非此即彼”关系为“亦此亦彼”关系；承认论域上存在并非完全属于该集合的元素，使经典结合的绝对属于变为相对属于的概念。

**定义 1:** 给定论域  $U$ ，所谓  $U$  上的一个模糊子集  $\tilde{A}$  是指，对于任意的  $x \in U$ ，都能确定  $\mu_{\tilde{A}} \in [0,1]$ ，用这个数表示  $x$  属于  $\tilde{A}$  的程度。映射：

$$\begin{aligned}\mu_{\tilde{A}}: U &\rightarrow [0,1] \\ x &\rightarrow \mu_{\tilde{A}} \in [0,1]\end{aligned}\tag{2}$$

称为  $\tilde{A}$  的隶属度函数，常用  $\mu_{\tilde{A}}(x)$  叫做  $U$  中的元素对模糊子集  $\tilde{A}$  的隶属度。隶属度  $\mu_{\tilde{A}}$  表示  $x$  属于  $\tilde{A}$  的程度， $\mu_{\tilde{A}}(x)$  越接近于 0，表示  $x$  隶属于  $\tilde{A}$  的程度越小； $\mu_{\tilde{A}}(x)$  越接近于 1，表示  $x$  隶属于  $\tilde{A}$  的程度越大；若  $\mu_{\tilde{A}}(x)$  越接近于 0.5，则表示  $x$  隶属于模糊集合  $\tilde{A}$  的程度越模糊。

## (2) 最大值和最小值贴近度

最大值和最小值贴近度  $N_M(\tilde{A}, \tilde{B})$  定义为：

$$N_M(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^n \mu_{\tilde{A}}(x_i) \wedge \mu_{\tilde{B}}(x_i)}{\sum_{i=1}^n \mu_{\tilde{A}}(x_i) \vee \mu_{\tilde{B}}(x_i)}\tag{3}$$

其中， $\mu_{\tilde{A}}(x_i)$ 、 $\mu_{\tilde{B}}(x_i)$  为  $U$  中元素  $x_i$  对模糊子集  $\tilde{A}$  和  $\tilde{B}$  的隶属度。 $\wedge$ 、 $\vee$  分别表示取小和取大运算符。

## (3) 灰关联度简介<sup>[2]</sup>

灰色系统理论 (Grey Theory) 是由华中科技大学的邓聚龙教授于上世纪 70 年代末、80 年代初期提出，已广泛应用于社会、经济、科技、农业、生态、生物等各个领域。

灰色系统分析方法针对不同问题性质有几种不同做法，灰色关联度分析 (Grey Relational Analysis) 是其中的一种。基本上灰色关联度分析是依据各因素数列曲线形状的接近程度做发展态势的分析。

灰色系统理论提出了对各子系统进行灰色关联度分析的概念，意图透过一定的方法，去寻求系统中各子系统 (或因素) 之间的数值关系。简言之，灰色关联度分析的意义是指在系统发展过程中，如果两个因素变化的态势是一致的，即同步变化程度较高，则可以认为两者关联较大；反之，则两者关联度较小。因此，灰色关联度分析对于一个系统发展变化态势提供了量化的度量，非常适合动态 (Dynamic) 的历程分析。

灰色关联度可分成“局部性灰色关联度”与“整体性灰色关联度”两类。主要的差别在于局部性灰色关联度有一参考序列，而整体性灰色关联度



是任一序列均可作为参考序列。

(4) 灰关联度模型

灰关联度是在一定的范畴内测度的量，因此，首先介绍灰关联空间的概念。

灰关联空间的构造机制属于公理化体系，既先有公理，然后根据公理来构造灰关联度的比较测度。

在灰关联差异信息空间  $\Delta_{GR}$  中：

$$@_{GRF} \Rightarrow \Delta_{GR}$$

$$\Delta_{GR} = (\Delta, \varsigma, \Delta_{0i}(\max), \Delta_{0i}(\min)) \quad (4)$$

$$\Delta = \{\Delta_{0i}(k) | i \in I, k \in K\}, \text{ or } \Delta = \{\Delta_{0i}(k) | i \in I\} \quad (5)$$

$$\Delta_{0i}(k) = |x_0(k) - x_i(k)| \quad (6)$$

点与点之间的比较测度称为灰关联系数；序列与序列间的比较测度称为灰关联度。

灰关联空间实质上是满足灰关联公理的差异信息测度空间。

**定义 2：** 令  $X$  为灰关联因子集， $\Delta_{GR}$  为灰关联差异信息空间，令  $\Gamma$  为灰关联映射集，则称  $(X, \Gamma)$  为灰关联空间因子态，称  $(\Delta_{GR}, \Gamma)$  为灰关联空间差异信息态。 $(X, \Gamma)$  和  $(\Delta_{GR}, \Gamma)$  统称灰关联空间。

**定理 1：** 令  $(\Delta_{GR}, \Gamma)$  为灰关联空间差异信息态

$$\Delta_{GR} = (\Delta, \varsigma, \Delta_{0i}(\max), \Delta_{0i}(\min)), \quad (7)$$

$$\Delta = \{\Delta_{0i}(k) | i \in I, k \in K\}, \text{ or } \Delta = \{\Delta_{0i}(k) | i \in I\}, \quad (8)$$

$$\Delta_{0i}(k) = |x_0(k) - x_i(k)| \quad (9)$$

$\Gamma$  满足灰关联 4 公理，

则  $(\Delta_{GR}, \Gamma)$  下有灰关联系数  $\gamma(x_0(k), x_i(k))$  为

$$\gamma(x_0(k), x_i(k)) = \frac{\min_i \min_k \Delta_{0i}(k) + \varsigma \max_i \max_k \Delta_{0i}(k)}{\Delta_{0i}(k) + \varsigma \max_i \max_k \Delta_{0i}(k)} \quad (10)$$

$$\Delta_{0i}(k) = |x_0(k) - x_i(k)| \quad (11)$$

$$\varsigma \in [0, 1],$$

则灰关联度  $\gamma(x_0, x_i)$  为

$$\gamma(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n \gamma(x_0(k), x_i(k)) \quad (12)$$

#### 5.1.4 问题 1 模型 II 求解——最大值和最小值贴近度的计算

首先，利用原始数据(附录 1.1)，我们分别求出它们的隶属度函数。由于七个指标与城镇就业人数都是效益型指标，因此，考虑用如下隶属度函数：

$$\mu_{A_i}(x) = \frac{x}{\max(X_i)} \quad i = 1, 2, 3, \dots, 8 \quad (13)$$

$$\text{其中, } \mu_{\tilde{A}_i}(x) \text{ 表示 } \begin{cases} \text{指标 } X_i \text{ 的隶属度函数} & i=1,2,\dots,7 \\ Y \text{ 的隶属度函数} & i=8 \end{cases} \quad (14)$$

于是, 可以求得各指标与城镇就业人数的模糊子集, 如果表 2 所示;

表 2 各项指标与城镇就业人数的模糊子集

$\tilde{A}_1$	$\tilde{A}_2$	$\tilde{A}_3$	$\tilde{A}_4$	$\tilde{A}_5$	$\tilde{A}_6$	$\tilde{A}_7$	$\tilde{A}_8$
0.8609	0.0636	0.0732	0.0621	0.7607	0.8308	0.0322	0.5641
0.8721	0.0741	0.0801	0.0724	0.8123	0.8332	0.0407	0.5781
0.8823	0.0933	0.0928	0.0895	0.8381	0.8574	0.0535	0.5912
0.8924	0.1274	0.1153	0.1175	0.8132	0.9243	0.0734	0.6045
0.9025	0.1733	0.1553	0.1603	0.8095	1.0000	0.0988	0.6174
0.9120	0.2200	0.1882	0.2022	0.7925	0.9436	0.1279	0.6303
0.9216	0.2563	0.2125	0.2367	0.7903	0.8727	0.1601	0.6595
0.9309	0.2830	0.2214	0.2627	0.8241	0.8284	0.1915	0.6879
0.9395	0.2890	0.2559	0.2807	0.8737	0.7994	0.2199	0.7155
0.9472	0.3014	0.2855	0.2983	0.9109	0.7945	0.2523	0.7419
0.9544	0.3333	0.3206	0.3300	0.9410	0.8090	0.2833	0.7663
0.9610	0.3581	0.3719	0.3647	0.9756	0.8114	0.3332	0.7925
0.9673	0.3860	0.4250	0.4002	1.0000	0.7994	0.3894	0.8203
0.9731	0.4430	0.4803	0.4517	0.9944	0.8155	0.4656	0.8487
0.9788	0.5204	0.5482	0.5317	0.9738	0.8372	0.5348	0.8764
0.9846	0.6112	0.6283	0.6094	0.9665	0.8203	0.6287	0.9047
0.9898	0.7172	0.7185	0.7048	0.9641	0.8179	0.7273	0.9371
0.9949	0.8369	0.8530	0.8299	0.9669	0.8445	0.8491	0.9715
1.0000	1.0000	1.0000	1.0000	0.9664	0.8533	1.0000	1.0000

然后, 利用 MATLAB 7.0 编程 (附录 2.1), 即可求得七个指标与城镇就业人数的最大最小值贴近度如表 3 所示:

表 3 七个指标与城镇就业人数的最大最小贴近度

影响因素	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
城镇就业人员							
$Y$	0.8009	0.4954	0.491	0.4896	0.8388	0.8244	0.4516

由表 3 可知, 在影响城镇就业人数的 7 个因素中,  $X_5$  对  $Y$  的最大最小贴近度最大, 因此说明指标  $X_5$  与  $Y$  最贴近,  $X_7$ 、 $X_4$  与  $Y$  的最大最小贴近度最小和次小, 说明  $X_7$ 、 $X_4$  的与  $Y$  距离最远。

### 5.1.5 问题 1 模型 II 求解——灰关联度计算

由灰色关联度的模型, 利用 MATLAB7.0 软件编程 (附录 2.1), 我们可以求得影响就业的七个因素与城镇就业人数的灰色关联度如表 4 所示:

表 4 影响就业的七个指标与城镇就业人员的灰色关联度

影响因素 城镇就业人数	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
$Y$	0.9839	0.7902	0.8219	0.7891	0.9886	0.978	0.6845

由表 4 可知：

$$\gamma(X_5, Y) > \gamma(X_1, Y) > \gamma(X_6, Y) > \gamma(X_3, Y) > \gamma(X_2, Y) > \gamma(X_4, Y) > \gamma(X_7, Y) \quad (15)$$

$$X_5 \succ X_1 \succ X_6 \succ X_3 \succ X_2 \succ X_4 \succ X_7 \quad (16)$$

其中  $\succ$  表示偏序关系。

由 (15)、(16) 两式可知， $X_5$ 、 $X_1$  对  $Y$  的影响位于前两位， $X_7$ 、 $X_4$  对  $Y$  的影响位于最后两位。

### 5.1.6 问题 1 结论

问题 1 模型 I 通过逐步回归法，得出模型中影响  $Y$  的主要因素为 5 个，即： $X_1$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$ 。

在问题 1 模型 II 中，综合最大最小贴近度和灰色关联分析发现，这两种方法得出的结论基本一致，而且结论都是  $X_7$  与  $X_4$  这两个指标排在最后两位，因此，可以认为它们两对  $Y$  的影响相对较小。由此我们可以得出影响  $Y$  的主要因素为： $X_5$ 、 $X_1$ 、 $X_6$ 、 $X_3$ 、 $X_2$ 。

## 5.2 问题 2 模型与求解

相对于问题 1 的二个模型，对于问题 2 我们也有二个模型。

### 5.2.1 问题 2 模型 I 的建立

问题 2 模型 I 是基于问题 1 模型 I 的结论上的。

由问题 1 模型 I 的结论可知，所选择的就业主要经济指标有：全国总人口 ( $X_1$ )、经济增长 GDP ( $X_4$ )、第三产业增加值 ( $X_5$ )、通货膨胀率 CPI ( $X_6$ )、货币政策  $M_2$  ( $X_7$ )。

利用 Eviews 6.0 进行线性回归，可得如下结果：

$$Y = -22517.59 + 0.364229X_1 - 0.054975X_4 + 0.063066X_5 - 20.01506X_6 + 0.032607X_7$$

$$t = (-13.41141) \quad (29.26797) \quad (-7.844341) \quad (2.158126) \quad (-4.335168) \quad (7.354697)$$

$$\bar{R} = 0.999530 \quad \bar{R}^2 = 0.999350 \quad F = 5535.080 \quad D.W = 2.335674$$

从上式中可以看出，各变量的统计意义都通过检验，但是从时序数列的角度来看，必须要确定各经济变量之间是否存在长期稳定的经济关系，才能进行计量分析。此外，从经济学角度考虑，多元线性的回归模型存在经济意义上的不足。基于此，我们从生产函数角度和内生经济增长理论角度对城镇就业问题模型进行理论分析。

#### 1. 从生产函数<sup>[3]</sup>的角度

经济学中有很多生产函数的形式，从复杂的 CES 生产函数到简单的完全替代生产函数，但是最为经典生产函数当属 C—D 生产函数，它描述的是劳动力与资本作为投入与产出之间的关系，其基本关系式为  $f = f(K, L) = Ae^{\lambda t} K^\alpha L^{1-\alpha}$ 。这一关系描述的是微观主体的投入产出关系，但同时也是我们宏观投入产出的基础。

基于此我们考虑城镇就业人数作为劳动力的投入与经济增长、产值、通胀等因素的关系也不是一个简单的线性关系，而是一种非线性关系，因此，我们需要考虑长期稳定的关系情况下，做非线性模型到线性模型的转化。

## 2. 从内生经济增长理论<sup>[4]</sup>的角度

内生经济增长理论也是描述产出与劳动力、资本以及技术等因素之间的关系，且产出与各经济因素之间也不是线性关系，从这个角度看，考虑就业这个问题，势必就是以劳动力作为基石来看，也就是说从内生经济增长理论来看，就业模型是一种非线性的关系。

基于上述两个角度的分析，对所得的数据(附录 1.1)进行协整分析，利用单位根检验，采用 ADF 指标进行判断。得出如下结果：

表 5 时序序列单位根检验结果

变量	检验形式	ADF 检验值	1%临界值	5%临界值	10%临界值	是否平稳
$\ln(X_1)$	$(c, t, p)$	-3.149553	-3.920350	-3.065585	-2.673459	是
$\ln(X_4)$	$(c, t, p)$	-3.766200	-4.728363	-3.759743	-3.324976	是
$\ln(X_5)$	$(c, t, p)$	-3.597098	-4.616209	-3.710482	-3.297799	是
$\ln(X_6)$	$(c, t, p)$	-3.490006	-4.616209	-3.710482	-3.297799	是
$\ln(X_7)$	$(c, t, p)$	-7.155931	-4.667883	-3.733200	-3.310349	是
$\ln(Y)$	$(c, t, p)$	-3.511996	-4.616209	-3.710482	-3.297799	是

注：检验类型括号中的  $c$ 、 $t$  和  $p$  分别表示检验平稳时估计方程的常数项单位根检验方程的常数项、时间趋势项和滞后阶数，0 表示不舍时间趋势项（下同）。

对各变量的协整检验表明： $\ln(X_1)$ 、 $\ln(X_4)$ 、 $\ln(X_5)$ 、 $\ln(X_6)$ 、 $\ln(X_7)$ 、 $\ln(Y)$  都是平稳数列。

基于此构造以下模型：

$$\ln(Y) = \beta_0 + \beta_1 \cdot \ln(X_1) + \beta_2 \cdot \ln(X_4) + \beta_3 \cdot \ln(X_5) + \beta_4 \cdot \ln(X_6) + \beta_5 \cdot \ln(X_7) + \varepsilon \quad (18)$$

利用 Eviews（见附录 3.1 的 var021）可得如下结果：

$$\begin{aligned} \ln(Y) = & 231.0035 - 19.46123\ln(X_1) - 1.25959\ln(X_4) + 1.361456\ln(X_5) \\ & - 0.496943\ln(X_6) + 0.854508\ln(X_7) \end{aligned} \quad (19)$$

$$t = (3.730631) \quad (-3.578484) \quad (-4.204063) \quad (3.908092)$$

$$(-3.104053) \quad (2.328970)$$

$$\bar{R} = 0.986892 \quad \bar{R}^2 = 0.981851 \quad F = 195.7543 \quad D.W = 1.122497$$

由（19）式，我们可以看出：

（1）全国总人口（ $X_1$ ）与城镇就业人数成负相关， $X_1$  每增加一个百分点时，将会导致城镇就业人数下降 19.46123 个百分点。

（2）GDP（ $X_4$ ）与城镇就业人数成负相关， $X_4$  每增加一个百分点，将会导致城镇就业人数下降 1.25959 个百分点。

（3）第三产业产值（ $X_5$ ）与城镇就业人数正相关， $X_5$  每增加一个百分点，

会导致城镇就业人数上升 1.361456 个百分点。

(4) CPI (  $X_6$  ) 与城镇就业人数成负相关,  $X_6$  每增加一个百分点时, 将会导致城镇就业人数下降 0.496943 个百分点。

(5) 货币政策 (  $X_7$  ) 与城镇就业人数正相关,  $X_7$  每增加一个百分点时, 将会导致城镇就业人数上升 0.854508 个百分点。

为了体现经济变量的滞后效应, 明确经济变量之间的内在关系, 我们建立 VAR 模型。通过对 LR、FPE、AIC、SC、HQ 这 5 个评价指标的分析, 确定 VAR 模型的最佳滞后期为 1。

得出问题 1 模型 I 的 VAR (1) 模型:

$$\ln[Y(t)] = -0.097308\ln[X_1(t-1)] + 0.060826\ln[X_4(t-1)] - 0.036023\ln[X_5(t-1)] - 0.069374\ln[X_6(t-1)] + 0.008156\ln[X_7(t-1)] + 0.896091\ln[Y(t-1)] + 2.125703$$

$$t = \begin{matrix} (-0.06308) & (0.69571) \\ (-0.30926) & (-1.61752) & (0.10408) \\ (18.0134) & (0.11892) \end{matrix}$$

$$\bar{R} = 0.999594 \quad \overline{R^2} = 0.999373 \quad F = 4519.222$$

### 5.2.2 问题 2 模型 II 的建立

问题 2 模型 II 是基于问题 1 模型 II 的结论。  
由问题 1 模型 II 结论可知, 所选择的就业主要经济指标有: 全国总人口 (  $X_1$  )、经济增长 GDP (  $X_4$  )、第三产业增加值 (  $X_5$  )、通货膨胀率 CPI (  $X_6$  )、货币政策  $M_2$  (  $X_7$  )。

利用 Eviews 6.0 软件 (见附录 3.1 的 eq022), 对  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_5$ 、 $X_6$ 、 $Y$  进行回归。

**Step1:** 对处理前的各指标作单位根检验  
在做回归模型之前, 我们要对 5 个自变量进行单位根检验, 观察它们的是否是平稳时间序列。5 个指标的 ADF 检验结果如图 1 至图 5 所示:

Null Hypothesis: X1 has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-21.32570	0.0000
Test critical values: 1% level	-3.857386	
5% level	-3.040391	
10% level	-2.660551	

图 1  $X_1$  的单位根检验结果

Null Hypothesis: X2 has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	7.396745	1.0000
Test critical values: 1% level	-3.857386	
5% level	-3.040391	
10% level	-2.660551	

图 2  $X_2$  的单位根检验结果

Null Hypothesis: X3 has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	12.94398	1.0000
Test critical values: 1% level	-3.857386	
5% level	-3.040391	
10% level	-2.660551	

Null Hypothesis: X5 has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	11.91806	1.0000
Test critical values: 1% level	-3.857386	
5% level	-3.040391	
10% level	-2.660551	

图 3  $X_3$  的单位根检验结果

Null Hypothesis: $X_3$ has a unit root Exogenous: Constant Lag Length: 0 (Automatic based on SIC, MAXLAG=0)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.445519	0.5371
Test critical values: 1% level	-3.857386	
5% level	-3.040391	
10% level	-2.660551	

图 4  $X_5$  的单位根检验结果

Null Hypothesis: $Y$ has a unit root Exogenous: Constant, Linear Trend Lag Length: 0 (Automatic based on SIC, MAXLAG=2)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.514368	0.3180
Test critical values: 1% level	-4.571559	
5% level	-3.690814	
10% level	-3.286909	

图 5  $X_6$  的单位根检验结果图 6  $Y$  的单位根检验结果

由图 1 我们可以知道，由于检验  $t$  统计量是-21.3257，远远小于显著性水平为 1%的临界值，至少表明可以在 99%的置信水平下拒绝原假设，认为  $X_1$  不存在单位根，是平稳的。但是，图 6 中的  $t$  统计量是-2.51，大于 10%的临界值，因此  $Y$  是非平稳的。同时，图 2 至图 5 表明，检验  $t$  统计量的值均远大于 10%的临界值，所以，不能拒绝原假设，即认为序列存在单位根，是非平稳的。

**Step2:** 对处理后的各指标做 ADF 检验

由 Step1 结论可知，在做回归模型之前，我们首先要对 5 个自变量和  $Y$  做变换，使得它们满足序列平稳性要求。经济学中有很多生产函数的形式，从复杂的 CES 生产函数到简单的完全替代生产函数，但是最为经典生产函数当属 C—D 生产函数，它描述的是劳动力与资本作为投入与产出之间的关系，其基本关系式为  $f = f(K, L) = Ae^{\lambda t} K^\alpha L^{1-\alpha}$ ，两边同时取对数，于是可以得到：

$$\ln f = \alpha \ln K + (1 - \alpha) \ln L + \ln A + \lambda t \quad (21)$$

由 (21) 式作变换，我们可以得到：

$$\ln L = \gamma \ln f - \alpha \gamma \ln K - \gamma \ln A - \gamma \lambda t \quad (22)$$

由 (22) 式，我们可以得知：劳动力投入取过对数后的指标与取过对数后的产出指标、资本投入指标成线性关系。因此，由此启发，我们对影响城镇就业人数以及影响它的 5 个指标取对数，然后再作回归分析。

对取过对数的各指标做 ADF 检验，结果如表 6 所示：

表 6  $\ln(Y)$  及其主要影响因素的 ADF 检验

变量	检验形式	ADF 检验值	1%临界值	5%临界值	10%临界值	是否平稳
$\ln(X_1)$	$(c, t, p)$	-26.11	-3.86	-3.04	-2.66	是
$\ln(X_2)$	$(c, t, p)$	-6.62	-4.62	-3.71	-3.30	是
$\ln(X_3)$	$(c, t, p)$	-3.94	-4.62	-3.71	-3.30	是

$\ln(X_5)$	$(c, t, p)$	-3.47	-3.86	-3.04	-2.66	是
$\ln(X_6)$	$(c, t, p)$	-1.41	-3.86	-3.04	-2.66	是
$\ln(Y)$	$(c, t, p)$	-3.51	-4.62	-3.71	-3.30	是

由表 6 可知道, 经过对数变换后,  $\ln(X_1)$ 、 $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_5)$ 、 $\ln(X_6)$ 、 $\ln(Y)$  均通过了单位根检验, 均为时间平稳数列。

**Step3:** 以  $\ln(Y)$  为被解释变量, 以  $\ln(X_1)$ 、 $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_5)$ 、 $\ln(X_6)$  为解释变量, 进行逐步回归, 结果如表 7 所示:

表 7  $\ln(Y)$  预测模型逐步回归结果

模型	$\ln(X_1)$	$\ln(X_2)$	$\ln(X_3)$	$\ln(X_5)$	$\ln(X_6)$	$R^2$	$\overline{R}^2$
$\ln(Y) = f(\ln(X_3))$			0.23 (26.07)			0.98	0.97
$\ln(Y) = f(\ln(X_3), \ln(X_1))$	0.40 (0.32)		0.21 (2.84)			0.98	0.97
$\ln(Y) = f(\ln(X_3), \ln(X_2))$		-0.22 (-5.21)	0.45 (10.60)			0.99	0.99
$\ln(Y) = f(\ln(X_3), \ln(X_5))$			0.45 (4.04)	-0.19 (-1.95)		0.98	0.98
$\ln(Y) = f(\ln(X_3), \ln(X_6))$			0.22 (28.60)		-0.31 (-3.01)	0.98	0.98
$\ln(Y) = f(\ln(X_3), \ln(X_2), \ln(X_1))$	0.71 (0.92)	-0.23 (-5.24)	0.41 (6.94)			0.99	0.99
$\ln(Y) = f(\ln(X_3), \ln(X_2), \ln(X_5))$		-0.26 (-4.38)	0.40 (5.23)	0.08 (0.84)		0.99	0.99
$\ln(Y) = f(\ln(X_3), \ln(X_2), \ln(X_6))$		-0.19 (-4.47)	0.41 (9.75)		-0.18 (-2.30)	0.99	0.99
$\ln(Y) = f(\ln(X_3), \ln(X_2), \ln(X_6), \ln(X_1))$	-0.55 (-0.60)	-0.18 (-3.77)	0.43 (7.90)		-0.21 (-2.09)	0.99	0.99
$\ln(Y) = f(\ln(X_3), \ln(X_2), \ln(X_6), \ln(X_5))$		-0.17 (-2.47)	0.43 (6.06)	-0.03 (-0.32)	-0.19 (-2.05)	0.99	0.99

由表 7 可知, 最终的回归模型为  $\ln(Y)$  与  $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_6)$  的线性回归, 回归结果如图 7 所示:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.705139	0.375862	23.16044	0.0000
LN <sub>X3</sub>	0.409710	0.042042	9.745159	0.0000
LN <sub>X2</sub>	-0.186388	0.041727	-4.466862	0.0005
LN <sub>X6</sub>	-0.172330	0.074958	-2.299019	0.0363
R-squared	0.993315	Mean dependent var	10.01572	
Adjusted R-squared	0.991978	S.D. dependent var	0.186939	
S.E. of regression	0.016743	Akaike info criterion	-5.156986	
Sum squared resid	0.004205	Schwarz criterion	-4.958157	
Log likelihood	52.99137	Hannan-Quinn criter.	-5.123337	
F-statistic	742.9518	Durbin-Watson stat	1.048460	
Prob(F-statistic)	0.000000			

图 7  $\ln(Y)$  与  $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_6)$  的回归模型

由图 7 我们可以得到相应的回归模型为：

$$\ln(Y) = 8.7051 - 0.1864\ln(X_2) + 0.4097\ln(X_3) - 0.1723\ln(X_6)$$

$$t = \quad \quad \quad (-4.47) \quad \quad (9.75) \quad \quad (2.30) \quad \quad (23)$$

$$R^2 = 0.9933 \quad \quad \quad F = 742.95 \quad \quad \quad D.W. = 1.0485$$

由 (23) 式可以看出，各解释变量的  $t$  统计量的绝对值大于 2.3，因此是显著的。可决系数  $R^2 = 0.9933$ ，说明该模型具有很高的拟合优度， $F$  检验也是高度显著的，回归系数的符号和数值是合理的。

接下来，我们要探讨  $\ln(Y)$  与  $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_6)$  是否存在长期的均衡关系，即要对这四个指标进行协整检验。

#### Step4: 协整检验 (Johansen 检验)

Johansen 于 1988 年，以及与 Juselius 一起于 1990 年提出了一种基于向量自回归模型的多重协整检验方法，通常称为 Johansen 检验，或 JJ 检验，是一种进行多重协整检验的较好方法。

利用 Eviews 6.0 (附录 3.1 的 eq022) 对  $\ln(Y)$  与  $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_6)$  做 Johansen 检验，得到结果如图 8 和图 9 表示：

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.916693	98.61711	47.85613	0.0000
At most 1 *	0.819022	56.36838	29.79707	0.0000
At most 2 *	0.699638	27.30889	15.49471	0.0005
At most 3 *	0.332115	6.861876	3.841466	0.0088

图 8 Johansen Cointegration Test 结果 1



Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.916693	42.24873	27.58434	0.0003
At most 1 *	0.819022	29.05949	21.13162	0.0031
At most 2 *	0.699638	20.44701	14.26460	0.0047
At most 3 *	0.332115	6.861876	3.841466	0.0088

图9 Johansen Cointegration Test 结果2

由图8和图9,我们不难发现,Johansen 检验结果是显著的,p 值均小于 0.05,因此,可以认为 $\ln(Y)$ 与 $\ln(X_2)$ 、 $\ln(X_3)$ 、 $\ln(X_6)$ 之间是协整的,因此(22)式是合理的。

#### Step5:异方差性检验

在 Eviews 6.0 软件上利用 White 检验对公式(22)作异方差检验,结果如图10 所示:

Heteroskedasticity Test: White

F-statistic	4.247649	Prob. F(9,9)	0.0212
Obs*R-squared	15.37933	Prob. Chi-Square(9)	0.0810
Scaled explained SS	5.038120	Prob. Chi-Square(9)	0.8310

图10 White 检验结果

由于 $nR^2 = 15.3793 < \chi_{0.05}^2(9) = 16.92$ ,所以,接受原假设,即认为公式(23)不存在异方差性。

#### Step6:自相关性检验

我们采用 D.W. 检验来验证公式(23)是否存在自相关性。由于 $n=19$ , $k=3$ ,取显著性水平 $\alpha=0.05$ ,查表可得 $d_L=0.97$ , $d_u=1.68$ ,而由图7可知, $D.W.=1.05$ ,则有 $d_L < D.W. < d_u$ ,因此,无法判断是否存在自相关性。

利用偏相关系数检验,则可得到结果如图11 所示:




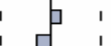

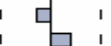







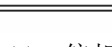
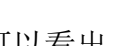
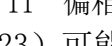
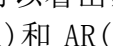
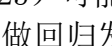






Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.450	0.450	4.4919	0.034
		2 -0.149	-0.441	5.0108	0.082
		3 -0.232	0.098	6.3544	0.096
		4 -0.123	-0.127	6.7539	0.149
		5 -0.117	-0.125	7.1459	0.210
		6 0.022	0.183	7.1604	0.306
		7 0.067	-0.177	7.3090	0.397
		8 -0.041	-0.019	7.3697	0.497
		9 -0.208	-0.210	9.0986	0.428
		10 -0.267	-0.182	12.262	0.268
		11 -0.188	-0.060	14.020	0.232
		12 -0.124	-0.292	14.898	0.247

图11 偏相关系数检验结果

由图11 可以看出,公式(23)可能存在一阶自相关和二阶自相关。但是,当我们把 AR(1)和 AR(2)加进去做回归发现, $t$  值不显著。因此,综合判断还是

决定用公式 (23) 作为最终的回归模型。

### 5.2.3 问题 2 模型 II 的结论

式 (23) 为:

$$\ln(Y) = 8.7051 - 0.1864\ln(X_2) + 0.4097\ln(X_3) - 0.1723\ln(X_6)$$

$$t = \quad \quad \quad (-4.47) \quad \quad \quad (9.75) \quad \quad \quad (2.30)$$

$$R^2 = 0.9933 \quad \quad \quad F = 742.95 \quad \quad \quad D.W. = 1.0485$$

由 (23) 式, 我们可以看出:

城镇就业人数与  $X_i$  ( $i=2, 3, 6$ ) 相关, 相关程度由各项的系数决定。

同样的, 可以确定问题 2 模型 II VAR 模型的最佳滞后期为 2, 确定建立 VAR(2) 模型, 可用 Eviews (见附录 3.1 的 var022) 得出结果如下:

$$\ln[Y(t-1)] = 2.01791 + 0.40051\ln[X_2(t-1)] - 0.035249\ln[X_2(t-2)]$$

$$+ 0.018521\ln[X_3(t-1)] + 0.02752\ln[X_3(t-2)] - 0.171633\ln[X_6(t-1)] \quad (24)$$

$$+ 0.0819375\ln[X_6(t-2)] + 0.49954\ln[Y(t-1)] + 0.299153\ln[Y(t-2)]$$

$$t = \quad \quad \quad (3.59179) \quad \quad \quad (2.00759) \quad \quad \quad (-1.76116)$$

$$\quad \quad \quad (0.87513) \quad \quad \quad (1.48357) \quad \quad \quad (-4.47380)$$

$$\quad \quad \quad (3.98784) \quad \quad \quad (3.99374) \quad \quad \quad (2.58643)$$

$$\bar{R} = 0.999960 \quad \quad \quad \bar{R}^2 = 0.999919 \quad \quad \quad F = 24791.96$$

### 5.3 问题 3 模型

针对本题, 我们考虑从分地区的角度入手, 尝试通过深入研究全国各地区的城镇就业人数 ( $Y$ ) 与各地区总人口 ( $X_1$ )、各地区经劳动生产率 ( $X_2$ )、各地区在岗职工工资 ( $X_3$ )、各地区 GDP ( $X_4$ )、各地区第三产业产值 ( $X_5$ )、各地区消费价格指数 CPI ( $X_6$ ) 的关系来确定它们之间的数学模型。

通过国家统计局网站、中国经济信息网、中国劳动力统计年鉴等, 我们整理出 1998 年到 2007 年间的全国 31 个省、市、自治区的城镇就业人数以及影响它的六个因素的数据, 数据见附录 1.2。

由于数据是时间序列数据与截面数据混合的面板数据, 因此, 首先在 Eviews6.0 软件中建立合并数据库, 即按指标把全国各地区 1998——2007 年数据合并成一行 310 行的数据。合并后各指标相应地定义为  $X_{1?}$ 、 $X_{2?}$ 、 $X_{3?}$ 、 $X_{4?}$ 、 $X_{5?}$ 、 $X_{6?}$ 。

对于合并后按指标分类的各组数据, 根据第二问的启示, 对每组数据取对数, 以期得到时间序列平稳和协整的效果。处理后的各组数据分别定义为  $\ln(X_{1?})$ 、 $\ln(X_{2?})$ 、 $\ln(X_{3?})$ 、 $\ln(X_{4?})$ 、 $\ln(X_{5?})$ 、 $\ln(X_{6?})$ 、 $\ln(Y?)$ 。

#### 5.3.1 问题 3 数据的检验

对经取对数处理后的各组数据, 分别进行单位根检验和协整检验。

- (1) 利用 Common root-Levin, Lin, Chu 检验方法对上述数据进行单位根检验, 所得结果如表 8 所示:

表 8 各项指标的 ADF 检验结果

指标	Method	Statistic	Prob.**	是否平稳
$\ln(X_{1?})$	Levin, Lin & Chu t*	-55.0568	0.0000	是
$\ln(X_{2?})$	Levin, Lin & Chu t*	-6.71769	0.0000	是
$\ln(X_{3?})$	Levin, Lin & Chu t*	-7.95638	0.0000	是
$\ln(X_{4?})$	Levin, Lin & Chu t*	-6.12139	0.0000	是
$\ln(X_{5?})$	Levin, Lin & Chu t*	-2.26471	0.0118	是
$\ln(X_{6?})$	Levin, Lin & Chu t*	-29.4487	0.0000	是
$\ln(Y_?)$	Levin, Lin & Chu t*	-9.97879	0.0000	是

由表 8 可知, 各项指标单位根检验的  $p$  值均小于 0.05, 因此, 接受原假设, 即认为各指标的序列是平稳的。

#### (2) 协整检验 (Johansen 检验)

对  $\ln(Y_?)$ 、 $\ln(X_{1?})$ 、 $\ln(X_{2?})$ 、 $\ln(X_{3?})$ 、 $\ln(X_{4?})$ 、 $\ln(X_{5?})$ 、 $\ln(X_{6?})$  进行 Johansen 检验, 所得结果如图 12 所示:

Newey-West bandwidth selection with Bartlett kernel				
Alternative hypothesis: common AR coefs. (within-dimension)				
	Statistic	Prob.	Weighted Statistic	Prob.
Panel v-Statistic	-7.043282	1.0000	-7.692278	1.0000
Panel rho-Statistic	5.468038	1.0000	6.482188	1.0000
Panel PP-Statistic	-17.05679	0.0000	-13.25879	0.0000
Panel ADF-Statistic	-7.288064	0.0000	-14.62410	0.0000
Alternative hypothesis: individual AR coefs. (between-dimension)				
	Statistic	Prob.		
Group rho-Statistic	7.906735	1.0000		
Group PP-Statistic	-26.98542	0.0000		
Group ADF-Statistic	-5.875939	0.0000		

图 12 Johansen 检验结果

由图 12, 根据 PP 统计量和 ADF 统计量, 拒绝非协整假设, 即各个序列是协整的。

#### 5.3.2 问题 3 模型选择

面板数据常用的三种模型为混合模型、个体固定效应模型和个体随机效应模型。在 Eviews 6.0 上分别做出三种模型, 然后再利用似然比检验和 Hausman 检验来选择最合适的模型。其中, 似然比检验是用来区分混合模型和个体固定效应模型; 而 Hausman 检验是用来区分个体固定效应模型和个体随机效应模型。具体检验结果如图 13 和图 14 所示:

Redundant Fixed Effects Tests			
Pool: POOL02			
Test cross-section fixed effects			
Effects Test	Statistic	d.f.	Prob.
Cross-section F	14.798766	(30,273)	0.0000
Cross-section Chi-square	299.321250	30	0.0000

图13 似然比检验结果

Correlated Random Effects - Hausman Test				
Pool: POOL02				
Test cross-section random effects				
<hr/>				
Test Summary	Chi-Sq. Statistic		Chi-Sq. d.f.	Prob.
<hr/>				
Cross-section random	47.451127		6	0.0000
<hr/>				
Cross-section random effects test comparisons:				
Variable	Fixed	Random	Var(Diff.)	Prob.
<hr/>				
LN1?	0.148480	0.094599	0.030356	0.7571
LN2?	-0.208278	-0.014783	0.001656	0.0000
LN3?	-0.320176	-0.542388	0.004642	0.0011
LN4?	0.875628	0.434809	0.007134	0.0000
LN5?	-0.095207	0.296706	0.004564	0.0000
LN6?	-0.370477	-0.437303	0.006692	0.4140
<hr/>				

图 14 Hausman 检验结果

由图13, 我们可以得知, 查  $F$  分布表, 可知在自由度为 (30, 273) 的前提下, 在5%的显著性水平下,  $F$  的临界值在1.54和1.57之间, 很明显, 此时有  $F > F_{0.05}(30, 273)$ 。所以, 拒绝原假设, 即认为个体固定效应模型要优于混合模型。

由图14, 自由度为6时, 在5%的显著性水平下的  $\chi^2$  分布的临界值为  $\chi_{0.05}^2(6) = 12.59$ , 此时  $H = 47.45 > \chi_{0.05}^2(6) = 12.59$ , 因此, 拒绝原假设, 即认为个体固定效应模型优于个体随机效应模型。

因此, 综合似然比检验和Hausman检验, 我们最终选用个体固定效应模型。利用Eviews 6.0软件(见附录3.2的model03), 我们得出个体固定效应回归结果如图15所示:

Dependent Variable: LNY?				
Method: Pooled Least Squares				
Date: 09/21/09 Time: 10:19				
Sample: 1998 2007				
Included observations: 10				
Cross-sections included: 31				
Total pool (balanced) observations: 310				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.008047	2.325503	2.153533	0.0322
LN1?	0.148480	0.192500	0.771323	0.4412
LN2?	-0.208278	0.060263	-3.456154	0.0006
LN3?	-0.320176	0.083499	-3.834506	0.0002
LN4?	0.875628	0.174307	5.023480	0.0000
LN5?	-0.095207	0.126382	-0.753330	0.4519
LN6?	-0.370477	0.414299	-0.894225	0.3720
Fixed Effects (Cross)				
BJ-C	0.730775			
TJ-C	0.271662			
HB-C	-0.339391			
SX-C	0.164034			
NMG-C	0.030973			
LN-C	0.269252			
JL-C	0.187263			
HLJ-C	0.222267			
SH-C	0.392311			
JS-C	-0.106560			
ZJ-C	-0.078427			
AH-C	-0.239299			
FJ-C	-0.169055			
JX-C	-0.209048			
SD-C	-0.142596			
HN-C	-0.270997			
HUB-C	-0.031136			
HUN-C	-0.214209			
GD-C	0.132902			
GX-C	-0.297044			
HAIN-C	-0.106601			
CQ-C	-0.166382			
SC-C	-0.229848			
GZ-C	-0.186574			
YN-C	-0.137917			
XZ-C	-0.282914			
SHX-C	0.143214			
GS-C	0.029910			
QH-C	0.154985			
NX-C	0.180167			
XJ-C	0.298282			
Effects Specification				
Cross-section fixed (dummy variables)				
R-squared	0.987645	Mean dependent var	6.014464	
Adjusted R-squared	0.986016	S.D. dependent var	0.850510	
S.E. of regression	0.100576	Akaike info criterion	-1.644188	
Sum squared resid	2.761564	Schwarz criterion	-1.198210	
Log likelihood	291.8492	Hannan-Quinn criter.	-1.465905	
F-statistic	606.2108	Durbin-Watson stat	0.998049	
Prob(F-statistic)	0.000000			

图15 个体固定效应回归结果

由图15，我们可以得出个体固定效应模型为：

$$\ln(Y_i) = (0.7308 + 5.0080)D_1 + (0.2717 + 5.0080)D_2 + \cdots + (0.2983 + 5.0080)D_{31} + 0.1485\ln(X_{1?}) - 0.2083\ln(X_{2?}) - 0.3202\ln(X_{3?}) + 0.8756\ln(X_4) - 0.0952\ln(X_5) - 0.3705\ln(X_6)$$

$$t = \quad (0.77) \quad (-3.46) \quad (-3.83) \quad (5.02) \quad (-0.75) \quad (-0.89)$$

$$R^2 = 0.9876 \quad \overline{R}^2 = 0.9860 \quad F = 606.2108$$

其中虚拟变量  $D_1, D_2, D_3, \dots, D_{31}$  的定义是：

$$D_i = \begin{cases} 1 & \text{如果属于第 } i \text{ 个个体}, i=1,2,\dots,31 \\ 0 & \text{其他} \end{cases}$$

### 5.3.3 模型优化

对个体固定效应模型进行优化，考虑到在六个因素情况下，存在多重共线性，采用逐步回归法，得出最终影响因素有：各地区经劳动生产率（ $X_2$ ）、各地区GDP（ $X_4$ ）、各地区第三产业产值（ $X_5$ ）、各地区消费价格指数CPI（ $X_6$ ）。再此基础上建立个体固定效应模型，得到结果如图16：

Dependent Variable: LNY?  
Method: Pooled Least Squares  
Date: 09/21/09 Time: 11:48  
Sample: 1998 2007  
Included observations: 10  
Cross-sections included: 31  
Total pool (balanced) observations: 310

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.798080	1.788291	3.801439	0.0002
LNx2?	-0.302654	0.047652	-6.351374	0.0000
LNx4?	0.956246	0.153121	6.245043	0.0000
LNx5?	-0.317353	0.113140	-2.804967	0.0054
LNx6?	-0.782585	0.409013	-1.913350	0.0567
Fixed Effects (Cross)				
BJ--C	0.656282			
TJ--C	0.145638			
HB--C	-0.068677			
SX--C	0.224305			
NMG--C	0.005957			
LN--C	0.494777			
JL--C	0.221159			
HLJ--C	0.377295			
SH--C	0.405678			
JS--C	0.253155			
ZJ--C	0.079155			
AH--C	-0.108011			
FJ--C	-0.048330			
JX--C	-0.160808			
SD--C	0.275837			
HN--C	0.012880			
HUB--C	0.214803			
HUN--C	-0.052030			
GD--C	0.476635			
GX--C	-0.283790			
HAIN--C	-0.491268			
CQ--C	-0.227215			
SC--C	-0.025992			
GZ--C	-0.321126			
YN--C	-0.150993			
XZ--C	-1.334633			
SHX--C	0.166164			
GS--C	-0.132837			
QH--C	-0.437007			
NX--C	-0.367272			
XJ--C	0.200269			
Effects Specification				
Cross-section fixed (dummy variables)				
R-squared	0.986963	Mean dependent var	6.014464	
Adjusted R-squared	0.985352	S.D. dependent var	0.850510	
S.E. of regression	0.102938	Akaike info criterion	-1.603384	
Sum squared resid	2.913937	Schwarz criterion	-1.181513	
Log likelihood	283.5245	Hannan-Quinn criter.	-1.434738	
F-statistic	612.3397	Durbin-Watson stat	0.935464	
Prob(F-statistic)	0.000000			

图16 个体固定效应模型结果



对模型结果进行异方差性及自相关性检验及修正，得到最终模型如图17：

Dependent Variable: LNY?					
Method: Pooled EGLS (Cross-section weights)					
Date: 09/21/09 Time: 11:51					
Sample (adjusted): 1999 2007					
Included observations: 9 after adjustments					
Cross-sections included: 31					
Total pool (balanced) observations: 279					
Iterate coefficients after one-step weighting matrix					
Cross-section weights (PCSE) standard errors & covariance (d.f. corrected)					
Convergence achieved after 19 total coef iterations					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	4.330754	0.483302	8.960760	0.0000	
LNx2?	-0.084755	0.030689	-2.761688	0.0062	
LNx4?	0.573873	0.082398	6.964660	0.0000	
LNx5?	-0.121414	0.055735	-2.178410	0.0303	
LNx6?	-0.308464	0.103584	-2.977902	0.0032	
AR(1)	0.798926	0.035779	22.32915	0.0000	
Fixed Effects (Cross)					
BJ-C	0.482825				
TJ-C	-0.397601				
HB-C	0.003975				
SX-C	0.077262				
NMG-C	-0.233893				
LN-C	0.438815				
JL-C	-0.083192				
HLJ-C	0.239979				
SH-C	0.152592				
JS-C	0.491795				
ZJ-C	0.454913				
AH-C	0.089155				
FJ-C	0.125783				
JX-C	-0.009142				
SD-C	0.458965				
HN-C	0.224889				
HUB-C	0.208058				
HUN-C	0.123201				
GD-C	0.728236				
GX-C	-0.059492				
HAIN-C	-0.571977				
CQ-C	-0.098377				
SC-C	0.290975				
GZ-C	-0.224960				
YN-C	0.032913				
XZ-C	-1.343069				
SHX-C	0.116747				
GS-C	-0.220837				
QH-C	-0.789969				
NX-C	-0.729912				
XJ-C	0.021343				
Effects Specification					
Cross-section fixed (dummy variables)					
Weighted Statistics					
R-squared	0.998912	Mean dependent var	11.98304		
Adjusted R-squared	0.998755	S.D. dependent var	6.957509		
S.E. of regression	0.076402	Sum squared resid	1.418452		
F-statistic	6372.734	Durbin-Watson stat	1.923828		
Prob(F-statistic)	0.000000				
Unweighted Statistics					
R-squared	0.990829	Mean dependent var	6.017409		
Sum squared resid	1.836098	Durbin-Watson stat	2.900167		

图17 异方差性及自相关性检验及修正的结果

### 5.3.4 问题3 结论

对于问题3，可以最终整理结果如下：

$$\ln(Y_t) = (0.482825 + 4.330754)D_1 + (-0.397601 + 4.330754)D_2 + \cdots + (0.021343 + 4.330754)D_{31} \quad (25)$$

$$- 0.084755\ln(X_2) + 0.573873\ln(X_4) - 0.121414\ln(X_5) - 0.308464\ln(X_6) + 0.798926AR(1)$$

$$t = \quad (-2.761688) \quad (6.964660) \quad (-2.178410) \quad (-2.977902) \quad (22.32915)$$

$$R^2 = 0.998912 \quad \overline{R^2} = 0.998755 \quad F = 6372.734$$

其中虚拟变量  $D_1, D_2, D_3, \dots, D_{31}$  的定义是:

$$D_i = \begin{cases} 1 & \text{如果属于第 } i \text{ 个个体}, i=1,2,\dots,31 \\ 0 & \text{其他} \end{cases}$$

同样由公式 (25) 可知道, 对全国总体而言:

$Y_i$  与  $X_i (i=2,4,6)$  的相关性与相关程度, 由  $X_i$  的系数决定。

#### 5.4 问题 4 仿真(预测)结果

利用第 2 个问题所建立的两套 VAR 模型, 对 2009 年以及 2010 年上半年我国城镇就业人员预测。

##### 5.4.1 预测方案 1

第一套预测方案, 所对应的 VAR 模型为公式 (20), 在 Eviews 软件上对样本内 (1990 年——2008 年) 和样本外 (2009 年——2010 年) 的城镇就业人数进行模拟和预测, 所得拟合图像如图 18——图 23 所示。图中蓝色为真实值, 红色为预测值 (下同)。

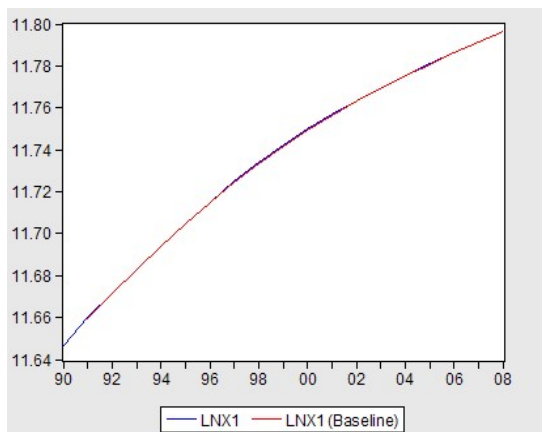


图 18 全国总人口的拟合图

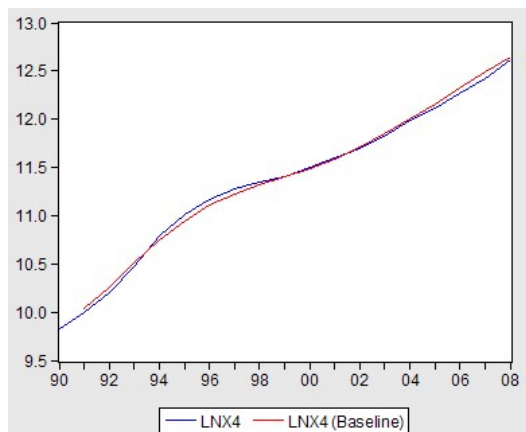


图 19 GDP 的拟合图

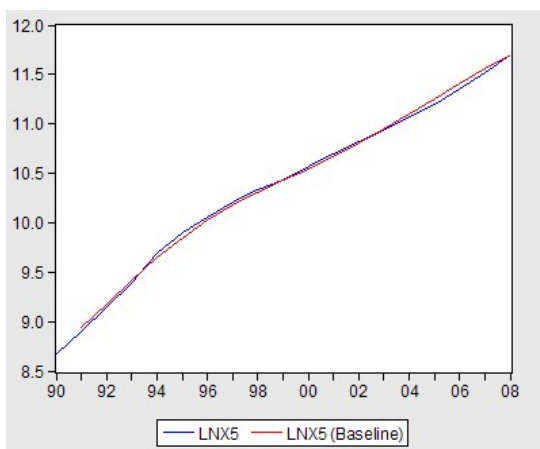


图 20 第三产业值的拟合图

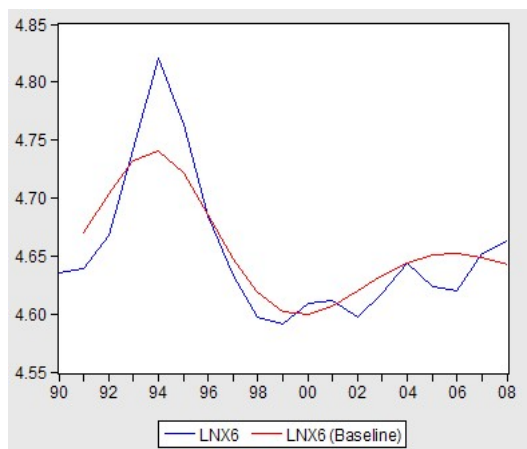


图 21 CPI 的拟合图

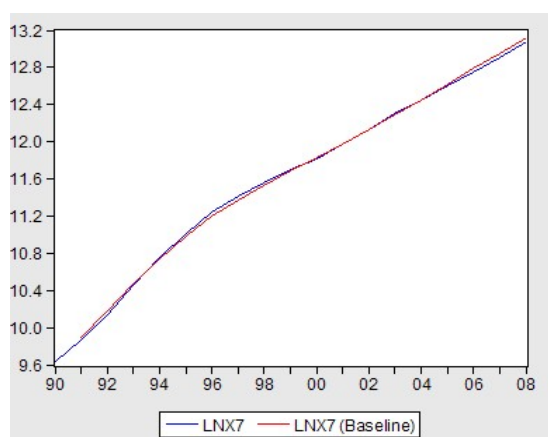


图 22 M2 的拟合图

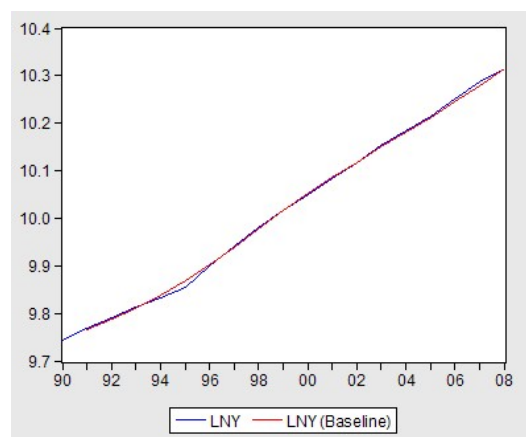


图 23 城镇就业人数拟合图

对于样本内预测的数据，由图 18——13 可以看出，除了 CPI 因素外，其余指标的预测精度都相当高。

样本外预测（2009 年和 2010 年）

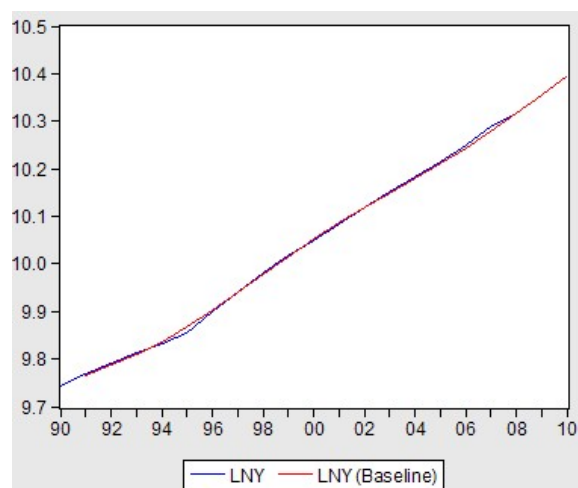


图 24 模型 1 城镇就业人数预测

由图，我们可以得出 2009 年与 2010 年的预测城镇就业人口：

$$\ln(Y_{2009}) = 10.35460 \Rightarrow Y_{2009} = 31401.1564 \text{ (万人)}$$

$$\ln(Y_{2010}) = 10.39501 \Rightarrow Y_{2010} = 32696.0645 \text{ (万人)}$$

由于 2008 年就业人口为 30210 万人，预测方案 1 所得 2009 年就业人口为 31401.1564 万人，两者差值为 1191.1564 万人，与中央预期 09 年新增就业 900 万人以上相符合。

#### 5.4.2 预测方案 2

第二套预测方案，所对应的 VAR 模型为公式 (24)，在 Eviews 软件上对样本内（1990 年——2008 年）和样本外（2009 年——2010 年）的城镇就业人数进行模拟和预测，所得拟合图像如图 25——图 28 所示。



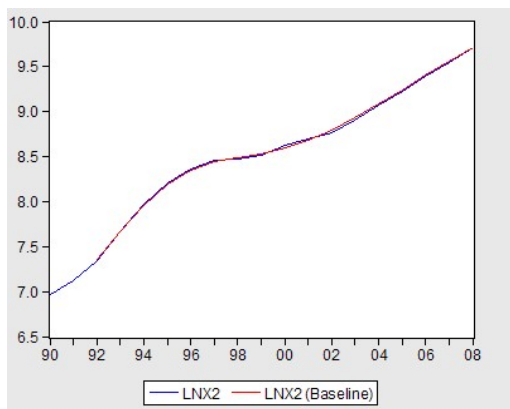


图 25 劳动生产率的拟合图

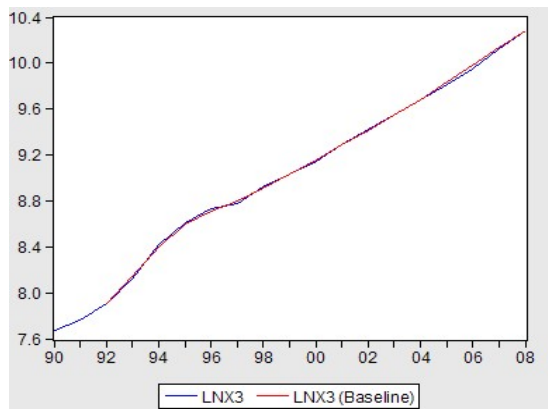


图 26 全国城镇单位平均工资拟合图

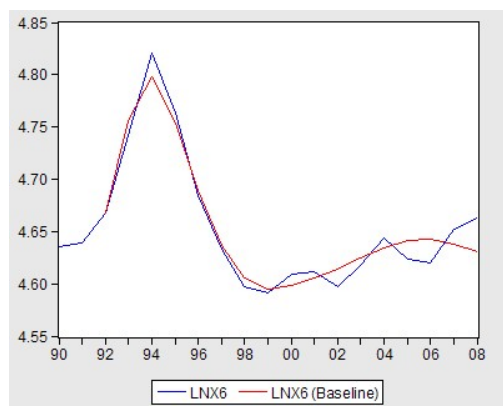


图 27 CPI 拟合图

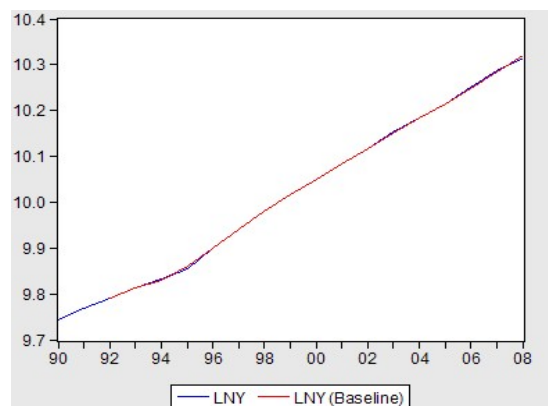


图 28 城镇就业人数拟合图

由图 25——图 28，同样是除了 CPI 指标外，其他几个指标的拟合程度都很好，说明该模型的拟合精度相当好。

样本外预测（2009 年和 2010 年）

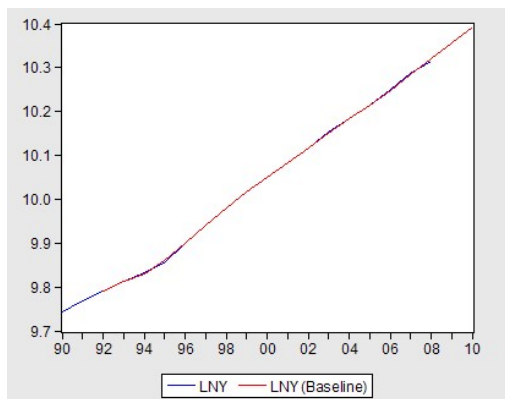


图 29 城镇就业人数预测

由图 29，我们可以预测出 2009 年以及 2010 年全国城镇就业人数，为：

$$\ln(Y_{2009}) = 10.35153 \Rightarrow Y_{2009} = 31304.9027(\text{万人})$$

$$\ln(Y_{2010}) = 10.39094 \Rightarrow Y_{2010} = 32563.26202(\text{万人})$$

由于 2008 年就业人口为 30210 万人，预测方案 1 所得 2009 年就业人口为

31304.9027 万人，两者差值为 1094.9027 万人，与中央预期 09 年新增就业 900 万人以上相符合。

#### 5.4.3 两个VAR模型优劣比较

一般认为，比较预测模型好坏的标准是样本期内的拟合效果，通常用误差预测值和真实值的误差平方和表示。我们在比较这两个模型优劣性时，除了考虑到样本期内的误差平方和以外，还充分考虑误差的变异系数和末期预测误差的大小。这样，评价才会更加合理。

由于两个 VAR 模型的滞后期不同，一个滞后 1 期，一个滞后 2 期，因此，我们考虑 1992 年——2008 年时期内的误差。我们整理出 1992 年——2008 年两个 VAR 模型的拟合误差值，如表 9 所示：

表 9 两种 VAR 模型预测精度比较

模型 年份	VAR1			VAR2		
	$\ln(Y)$	$\ln(Y_0)$	$R_1$	$\ln(Y)$	$\ln(Y_1)$	$R_2$
1992	9.790375	9.787306	-0.00307	9.790375	9.790217	-0.00016
1993	9.812578	9.810498	-0.00208	9.812578	9.813416	0.000838
1994	9.833762	9.836888	0.003126	9.833762	9.831048	-0.00271
1995	9.854297	9.867363	0.013066	9.854297	9.859396	0.005099
1996	9.89958	9.901837	0.002257	9.89958	9.899304	-0.00028
1997	9.941794	9.93917	-0.00262	9.941794	9.941187	-0.00061
1998	9.981189	9.977621	-0.00357	9.981189	9.980168	-0.00102
1999	10.01735	10.01551	-0.00184	10.01735	10.01643	-0.00092
2000	10.04979	10.05169	0.001897	10.04979	10.0507	0.000907
2001	10.08331	10.08576	0.002454	10.08331	10.08411	0.000804
2002	10.11779	10.118	0.000208	10.11779	10.11737	-0.00042
2003	10.15187	10.14911	-0.00276	10.15187	10.15034	-0.00153
2004	10.18399	10.18002	-0.00397	10.18399	10.18294	-0.00105
2005	10.21578	10.21159	-0.00419	10.21578	10.21567	-0.00011
2006	10.25097	10.24454	-0.00643	10.25097	10.24915	-0.00182
2007	10.28705	10.27928	-0.00777	10.28705	10.28371	-0.00334
2008	10.31593	10.31597	4.17E-05	10.31593	10.31928	0.003352

注意：VAR1 模型指的是公式（20），VAR1 模型指的是公式（24）， $\ln(Y)$ 是指 Y 原始数据取对数， $\ln(Y_0)$ 与  $\ln(Y_1)$ 分别指 VAR1 模型和 VAR2 模型的预测值， $R_1$  和  $R_2$  分别是两种预测方法的误差。以时间段为横坐标，以  $R_1$  和  $R_2$  为纵坐标，利用 MATLAB7.0 软件(程序见附录 2.2)在直角坐标系中作出两种预测方法产生误差的散点图，如图 30 所示：

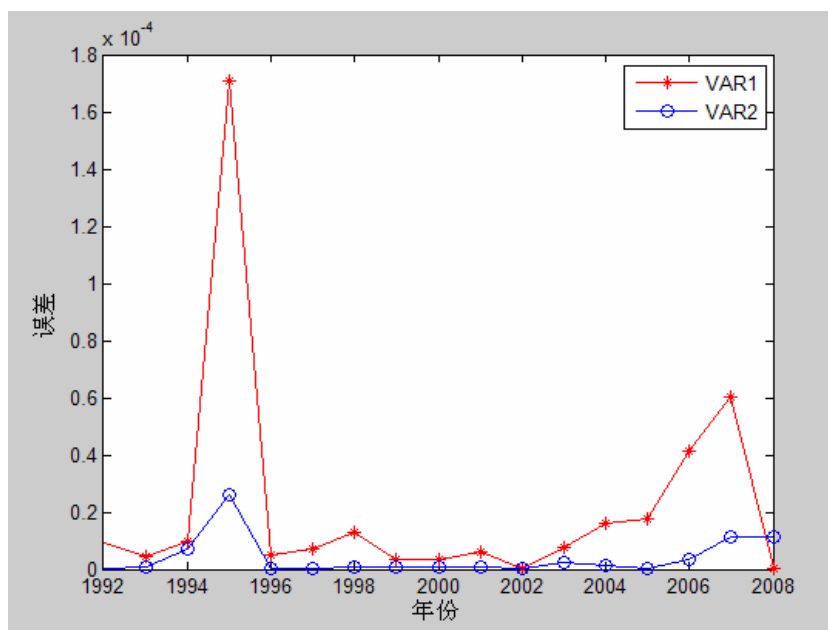


图 30 1992——2008 年 VAR1 与 VAR2 模型的误差散点图

由图 30 可知，VAR1 模型的误差变化很大，而且普遍比 VAR2 模型的误差要大，但是，末期（2008 年）的误差却比较小。

通过计算，则有  $S_1 = 1.8762 > S_2 = 1.7149$ ， $SR_1 = 0.3561 \times 10^{-7} > SR_2 = 0.01 \times 10^{-7}$

其中， $S_1$  与  $S_2$  分别是 VAR1 模型和 VAR2 模型误差的变异系数， $SR_1$ 、 $SR_2$  分别表示 VAR1 和 VAR2 的误差平方和。

因此，我们经过综合评价认为，VAR2 模型比 VAR1 模型预测效果好，只是末期 VAR1 模型预测效果稍微好些。

## 5.5 政策建议

### 5.5.1 模型结果及原因分析

根据问题 2 已有的模型分析过程，我们可以得出城镇就业人数与各变量之间的模型结果及其原因分析如下：

对于模型 1 来说，全国总人口、GDP、CPI 均与城镇就业人数呈反向变化，而三产增加值、M2 则与城镇就业人数呈正向变化。显然呈反向变化的变量似乎都不符合一般西方经济理论的分析，但是结合中国实际情况来看，其结果则有其必然性。这是因为目前中国的实际状况与经典西方经济理论存在较大的差异，奥肯定律和菲利普斯曲线的规律性在中国经济运行中都得不到体现，反而是恰恰相反，也就是说高失业率与高通胀率、高经济增长率同时存在。此外，我们看到人口的增长与就业人数也是成反向关系变化，主要在于中国人口“红利”将尽，劳动力所占的比重下降，所以总人口增加所能带来的城镇就业并不是一致增加的，当然这里面也有二元经济结构的“功劳”。

对于模型 2 来说，劳动生产率、CPI 与城镇就业人数呈反向变化，而城镇单位平均工资与城镇就业人数呈正向变化。显然这些变化在前面的分析中都表明是符合中国实际情况的。

根据问题 4 的仿真结果来看。两个模型的预测效果都相当不错，而且精度都是比较高，相对而言模型 2 的准确度更好一些，也就是表明在微观层面的因素影

响要好过于宏观调控上的影响程度，一定意义上也都说明了制度因素对于失业问题的顺利解决的重要性。因此，下面将从制度经济学的角度来分析提高城镇就业人数或减少城镇登记失业率的政策建议。

### 5.5.2 基于制度经济学角度的政策建议

#### 1. 失业问题的制度性

失业问题的产生无外乎有两个原因：一个就是基于凯恩斯理论考虑的有效需求不足；另一个则是源于技术和结构的变化。归结起来就是需求和供给两个方面的因素所导致，以就业问题来看，需求方面是有消费、投资、及货币等因素，而供给方面则有技术与结构等因素。无论是需求方面还是供给方面都与制度有着密不可分的关系，换言之，制度与总需求和总供给都是具有相关性，从这个角度来看，解决就业问题或失业问题其实质上就是一个制度问题。同时，考虑到我国现实表现出的二元结构特征，失业问题是具有很强的制度性：

（1）宏观经济规划目标中并没有把就业问题摆在优先位置上，单纯的追求经济增长看作发展的目标，政府在解决就业问题中主导作用没有体现。宏观经济规划是制度变迁过程中政府的选择，显而易见的是就业问题能否解决取决于政府的制度选择。

（2）在产业结构上也是有同样的制度缺陷，搞得失业率显示我们的二产与三产对就业的促进作用十分有限。基于此，考虑产业结构的转变也是解决就业问题的一个有效手段，亦是制度变迁过程中的合理选择。

#### 2. 制度对失业问题的影响

制度对于失业问题的影响表现在于各种体制和制度上，诸如经济体制、产权制度、社保制度等都会作用于经济因素进而作用到就业，例如对技术和结构的影响。总而言之，不同的制度环境会造就不同的经济发展战略，而这也都会影响到就业实际效果。

#### 3. 基于制度角度考虑就业问题

如前所分析结果，能否提高就业人数和降低失业率实际就是一个制度性问题能否得到有效的解决，也就是说需要从需求和供给两个方面来考虑政策建议。从需求方面来看，有效需求提升离不开消费、投资及货币等因素，因此为了解决失业问题就必须着手于以下几个方面：

（1）完善劳动力市场机制，建立有效工资机制，修正扭曲工资状况，形成弹性工资机制，确保充分就业的基础环境。

（2）有效的财政政策与货币政策配合。财政政策与货币政策都是通过刺激扩大总需求，进而增加就业机会。2008年中国经济面临重大考验时，政府果断采取4万亿的拉动内需投资计划就是一个很好的例证。但是需要注意的是在采取经济政策时应当避免直接给予，尽量创造劳动的机会进行支付，形成有效的工资机制。

（3）有效的出口贸易政策，出口贸易直接关联到就业问题。在现阶段来看，我国出口多数以劳动密集型为主，对就业问题的解决有着重大的影响。金融危机冲击下，我国出口有了大幅的下降，所以必须寻求潜在的进口国，开辟更多贸易国市场，增加就业机会。

新制度经济学中制度的起源在于交易费用，但是有效制度需求并不能得到相应的制度供给，制度的供给能否出现很大程度上取决于政府的作用。因此，从供给方面来看失业问题的解决势必要求政府作用发挥到最大，即需要着眼于以下几点：

(1) 产业结构调整与转换。目前来看,劳动力在一、二、三产中的分布比例并不是很合理,二产与三产对就业的贡献率并不好。因此,必须要调整好产业结构,重点在于二产向三产的转移,增加三产的比例,达到降低失业率的目的。

(2) 人力资源开发,提高劳动力素质。就业问题不能得到很好的解决的一个原因就是市场对劳动力的需求层次超过实际供给水平,存在“有工无人能做”和“有人无工可做”现象。开发优质人力资源,加强对需要就业的人员技能培训,做到供给与需求层次配合,提高就业人数。

(3) 完善产权制度。产权得不到明确势必会造成企业的发展规模和速度都会受到限制,进而影响到企业吸纳劳动力的能力。

(4) 发展合乎中国国情的技术创新。技术创新有助于私人部门缺陷的革除,引导其长期有效的发展,进而提升就业水平。另一方面,技术革新也会有助于公共部门提供有效的就业保障和更多就业机会。

(5) 完善社会保障制度。社保制度的完善是为失业人员暂时的保障,利于有效劳动力市场的形成,长期来看也是有利于提高就业人数的。

## 6 模型优缺点

### 优点:

1. 采用模糊贴近度与灰色关联度两种方法进行对比选择,得出结论可靠。
2. 对 31 各省市采取面板数据,综合情况分析,得到各个省份的就业模型,具有普遍意义。
3. 利用 VAR 动态预测方法进行预测,并进行模型对比,得到最优的预测结果且预测准确度较高。
4. 从制度经济学的角度对失业问题进行分析,并结合实际模型预测结果给出较为合理的政策建议。
5. 从不同角度考虑模型的选择,建立多个模型,并结合内生经济增长理论给出分析。

### 缺点:

1. 仅从年度数据和面板数据考虑,没有细化到月度数据,使得预测仅限于年度预测。
2. 由于时间关系,仅考虑到分地区条件下的模型建立,没有考虑分行业、分就业人群等情况下的模型建立。

## 7 参考文献

- [1] 杨纶标,高英仪,模糊数学原理及应用,华南理工大学出版社,135—140,2001
- [2] 邓聚龙,灰理论基础,武昌喻家山:华中科技大学出版社,2002
- [3] 高鸿业,吴易风,刘凤良,研究生用西方经济学,经济科学出版社,85—99,2006
- [4] 多恩布什,费希尔,斯塔兹,宏观经济学,中国财政经济出版社,66—74,2003
- [5] 袁乐平,周浩明,失业经济学,经济科学出版社,221—259,2003
- [6] 牛润霞,技术变迁中的失业问题研究,人民出版社,194—197,2007
- [7] 李子奈,潘文卿,计量经济学,高等教育出版社,322—362,2005
- [8] 蔡昉,中国二元经济与劳动力配置的跨世界调整——制度、结构与政治经

- 济学的考察, 浙江社会科学, 18—22, 2005
- [9] 周长文, 宏观调控与失业, 数量经济技术经济研究, 18—21, 2002
  - [10] 盛乐, 姚先国, 失业问题的微观分析与模型解释, 中国经济问题, 49—54, 2002
  - [11] 钱小英, 我们失业率的特征及其影响因素分析, 经济研究, 28—36, 1998
  - [12] 张怡, 有关中国失业问题的思想, 经济研究导刊, 116—117, 2009
  - [13] Aleksander Berentesen, Guido Menzio, Randall Wright, Inflation and unemployment in the long run, National Bureau of Economic Research, 2008.
  - [14] Blanchard, Olivier and Jordi Gali, A new Keynesian Model with Unemployment, Mimeo, MIT, 2005
  - [15] Mortensen, Dale T., Eva Nagypal, More on unemployment and Vacancy Fluctuations, Review of Economic Dynamics 10, 327—47. 2007